CHAPTER I

INTRODUCTION

Thalassemias are hereditary hemolytic anemias characterized by decreased or absent synthesis of one or more of the globin subunits of the hemoglobin molecule (1). These are usually classified according to the type of the globin chain which is absent or present in decreased amount. The alpha- and beta-thalassemias are the most common genetic disorders with a worldwide distribution, alpha-thalassemia alone may represents the most prevalent inherited disorder of mankind. It occurs widely through out Africa, the Mediteranean countries, the Middle East and Southeast Asia. The high gene frequency results from a selective advantage of the thalassemia of heterozygotes which are protected against severe malaria (2). In Thailand as many as 20% of the population are heterozygous for one or other type of thalassemia and alpha-thalassemia is much more common than beta-thalassemia (3). The frequencies are 20-30% of alpha-thalassemia, 3-9% of beta-thalassemia and at least 4% of hemoglobin Constant Spring (4,5).In the last decade the technique of DNA hybridization has been used to elucidate the molecular basis of the genetic abnormalities of thalassemias and other genetic disorders such as Duchennie muscular dystrophy, cystic fibrosis,Huntington's disease and others.In the present investigation, these methods were used to study the molecular defects and the frequencies of different alpha-thalassemia genes in Northern Thailand.

I.1 The structure of human hemoglobin

Human hemoglobin consists of two different pairs of identical peptide chains to which a heme group is attached. These chains are designated by Greek letters, alpha (α), beta (β), gamma (γ), delta (δ), zeta (ζ) and epsilon (ε). The normal molecule is made up of two alpha- (or α-like) and two non alpha-chains which differ in the developmental stages as shown below (3,6).

Embryonic Hemoglobins :

Hb Gower-1	$\zeta_2\varepsilon_2$	50%
Hb Gower-2	$\alpha_2\varepsilon_2$	25%
Hb Portland	$\zeta_2\gamma_2$	5%

Fetal Hemoglobins :

Hb F	$\alpha_2\gamma_2$	90%
Hb A	$\alpha_2\beta_2$	10%

Adult Hemoglobins :

Hb A	$\alpha_2\beta_2$	97%
Hb A_2	$\alpha_2\delta_2$	3%
Hb F	$\alpha_2\gamma_2$	<1%

In thalassemias, decreased or absent synthesis of globin chain results in a relative excess of the other chains. The excess chains are unstable and their precipitation causes damage to the cell membrane which results in changes of cell morphology and ineffective erythropoiesis (7). The alpha-thalassemias are most important in the fetal and newborn period, the beta-thalassemias in childhood and adult

life. Gamma-thalassemia would be of no importance in adults, gamma-beta-thalassemia has been responsible for neonatal hemolytic anemia which was ameliorated with maturation. Severe thalassemias involving gamma, epsilon or zeta loci may lead to fetal or embryonic death. Because of the proximity of beta- and delta-globin chain loci reduced synthesis of both beta- and delta-globin chains may occur and give rise to delta-beta-thalassemia or F-thalassemia. Delta-thalassemia has been rarely reported and is believed to be without clinical manifestation (8).

I.2 Recent methodologic advances in the study of genes and their Structure.

Many inherited diseases are known to run in families, the transmission following the rules of classic Mendelian of inheritance. Only one gene among the approximately 100,000 genes on the 23 pairs of human chromosomes is defective. The symptoms and progress of approximately 3,000 hereditary diseases have been described, but for most of them nothing or little is known about the biochemical basis. The last decade has witnessed an explosion of new knowledge about the organization and structure of many different gene systems and thalassemias are probably the group of diseases whose molecular basis is best known. (9). The chromosomal localization of the human globin genes has been determined, and the mutant genes leading to thalassemias have been studied in detail. The methods of " new genetics " have provided us with a greatly expanded set of markers : molecular variations known

as RFLP (Restriction Fragment Length Polymorphism) facilitate the examination of anomalous genes.

"New genetics" was made possible by a number of fundamental discoveries and technological developments in nucleic acid chemistry, microbial genetics and molecular biology, such as :

- The successful isolation of some mRNA.
- The possibility to produce globin cDNA by the use of reverse transcriptase enzyme.
- The possibility to prepare globin cDNA in radioactive form, that can be used as specific " probe" for the detection and quantification of globin mRNA or globin genes by molecular hybridization.
- The discovery of restriction endonuclease enzymes which digest DNA at specific polynucleotide sequences (palindromes).
- The successful incorporation of mammalian DNA sequences into plasmids or bacteriophages, which can replicate in bacterial hosts.
- The development of method for characterizing DNA-fragments which become to be known as " genomic blotting"

The RFLP defined a potential marker. A single restriction enzyme finds a large number of cutting sites in the total human DNA ($3x10^9$ bp per genome). The way to detect one or few variant fragments among millions is : The fragments are first sorted by electrophoresis, an electric field draws them throguh a gel. Then the Southern blotting

technique serves to detect the fragments of interest. A radioactive single stranded DNA that acts as a probe, detecting and binding to the complementary sequence in DNA sample that has been denatured (heated or exposed to the high pH in order to separate its strands), and blotted onto a membrane. The radioactive labelling makes it possible to detect the position of the fragments which reveal their size, and deleted gene fragments can be traced. This opens the way to simple tests for diagnosing carriers and future disease victims, particularly for the prenatal detection of potentially serious hemoglobin disorders.

Abnormalities caused by gene deletions and point mutations which alter a restriction enzyme recognition site can be detected by this technique, but not for the majority of point mutations. To detect these changes in DNA directly on the basis of hybridization, the size of the DNA probe used is critical. Theins and Wallace (10) developed a method of hybridization using 11,14 or 17 base long oligonucleotides. They observed that the duplex with a single base pair mismatch was significantly less stable than their perfectly matched counterparts on dissociation at a low temperature). This difference in thermal stability and the specific sequence in the region of the mutation is the basis of the oligonucleotide hybridization technique. It is now sucessfully used in the diagnosis of single base mutations.

The use of the genomic hybridization technique (Figure I.1), the oligonucleotide hybridization technique(Figure I.2) and the nucleotide

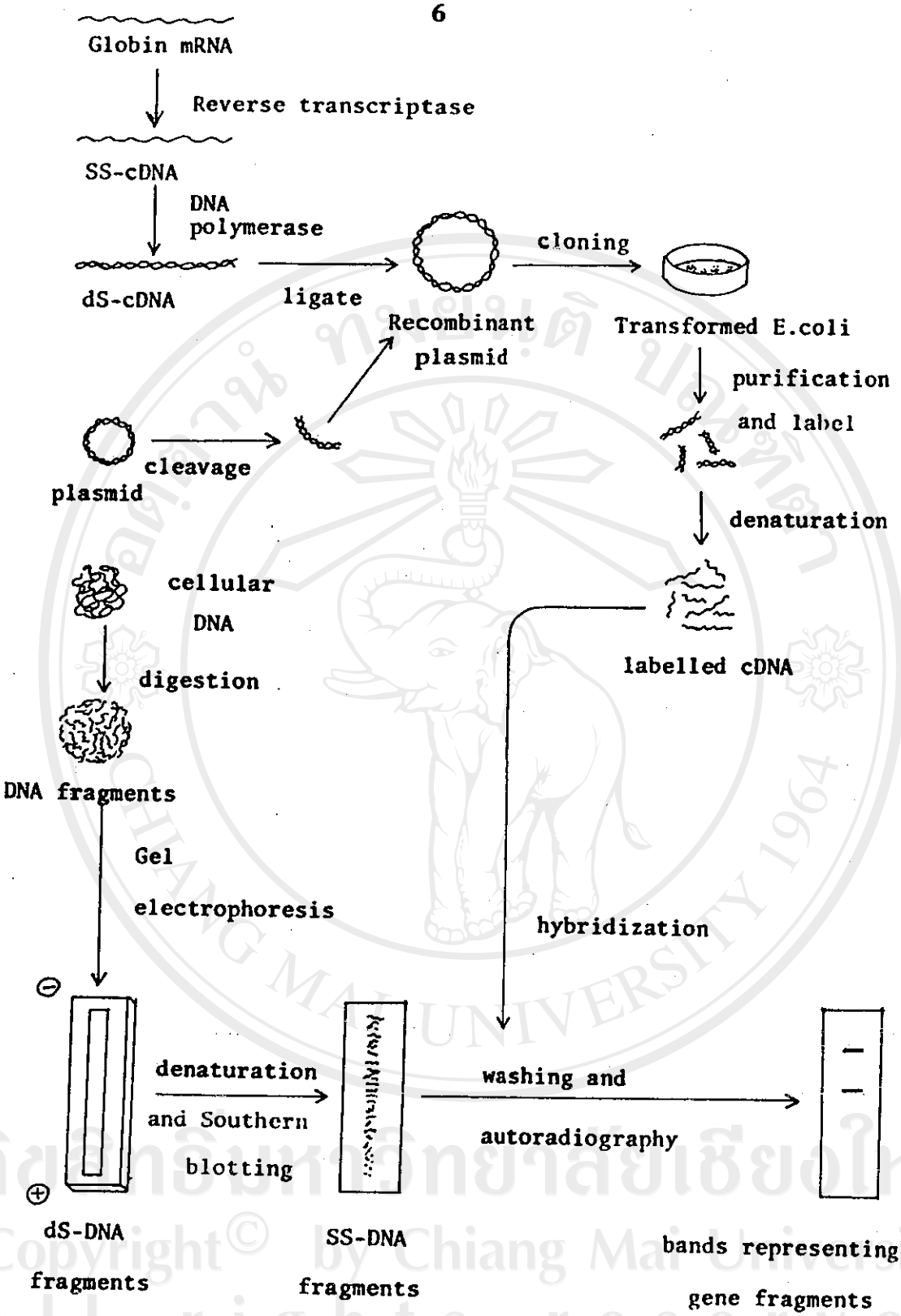

Figure I.1 The blotting hybridization technique for the demonstration of deletional alpha thalassemia detection.

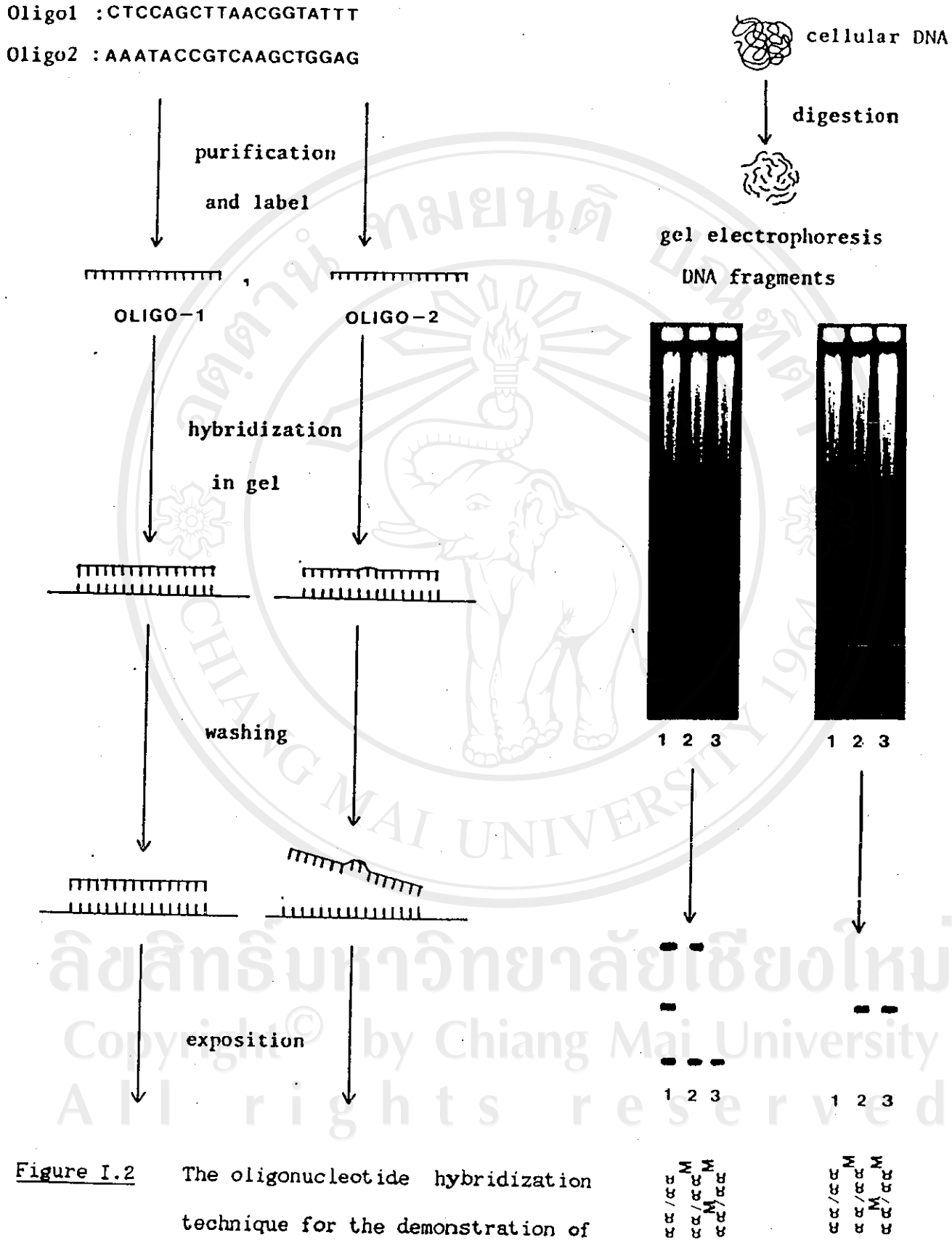

Figure I.2 The oligonucleotide hybridization technique for the demonstration of Hb Constant Spring detection.

base sequencing method made it possible to understand genetic mechanisms, the structure of globin genes and the abnormality of thalassemic genes.

I.3 The human globin gene structure :

The human hemoglobin subunits are encoded by two clusters of genes. The alpha globin gene cluster and the beta globin gene cluster.

The alpha globin gene cluster is found on the short arm of chromosome-16. It includes two alpha gene (α_1 and α_2), an embryonic alphalike gene (ζ), three pseudogenes ($\psi\zeta$, $\psi\alpha_1$ and $\psi\alpha_2$) and one gene of undeterminded status (θ), arrange in the order :

$$5' \text{——} \zeta - \psi\zeta - \psi\alpha_2 - \psi\alpha_1 - \alpha_2 - \alpha_1 - \theta \text{——} 3'$$

within a 40 kb segment (11, 12). The two linked alpha globin genes are indentical, both genes are transcriptionally active and encode identical protein products (11), but the $alpha_2$ gene is functionally the major alpha globin gene in humans, encoding two to three times as much alpha globin protein as the adjacent $alpha_1$ gene (13).

The beta globin gene cluster is found on the short arm of chromosome-11. It includes two gamma genes ($^{G}\gamma$ and $^{A}\gamma$ which differ in only one amino acid at codon 136 Gly→Ala), an embryonic gene (ε), two adult genes (δ and β) and a pseudogene ($\psi\beta$) arrange in the order :

$$5' \text{——} \varepsilon - {}^{G}\gamma - {}^{A}\gamma - \psi\beta - \delta - \beta \text{——} 3'$$

within a 50 kb segment (14,15).

The 5' to 3' sequence of the globin genes is in the same order that they are developmentally active : ζ to α and ε to β respectively. The significance of this organization in relation to the control of the switches in globin gene expression that occur during embryonic and fetal development is not known.

The globin genes include three coding blocks (exons) separated by two non-coding blocks (IVS = intervening sequences or introns) (9) (Figure I.3).

I.4 The genetic mechanism :

Studies using the genomic hybridization technique have shown that normal genes contain IVS that do not code for protein. Initially, the whole gene sequence is translated to form heterogeneous nuclear RNA (hnRNA) and the IVS are then removed by enzymatic cleavage and splicing give cytoplasmic mRNA. This is the template for protein synthesis (Figure I.4). The known defects in thalassemia are the point mutations (single base exchange),or gene deletion (6), for example :

a. Deletions of globin genes account for most alpha-thalassemias and for Hereditary Persistance of Hb F.
b. Partial deletions of the beta globin gene, preventing transcription to hnRNA occurs in some beta-thalassemias.
c. Abnormality of cleavage or splicing site in the IVS may prevent hnRNA processing to cytoplasmic mRNA.
d. Point mutations in the original structural gene may prevent transcription of the mRNA.

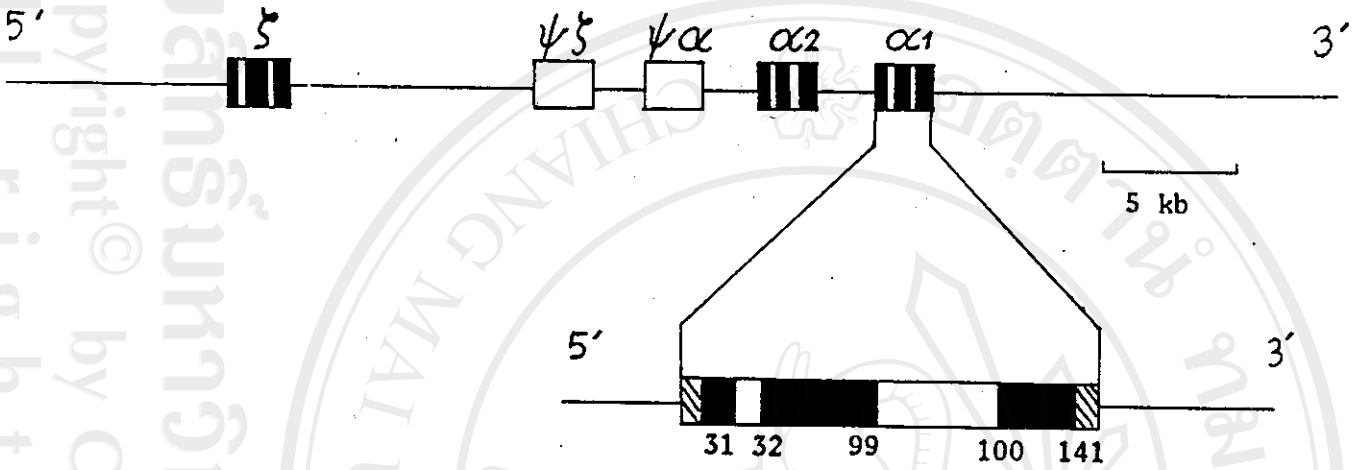

Figure I.3 The general pattern of organization of the human globin genes.

■ = the position of exons.

□ = the position of intervening sequences.

The amino acid residues corresponding to the codons are indicated.

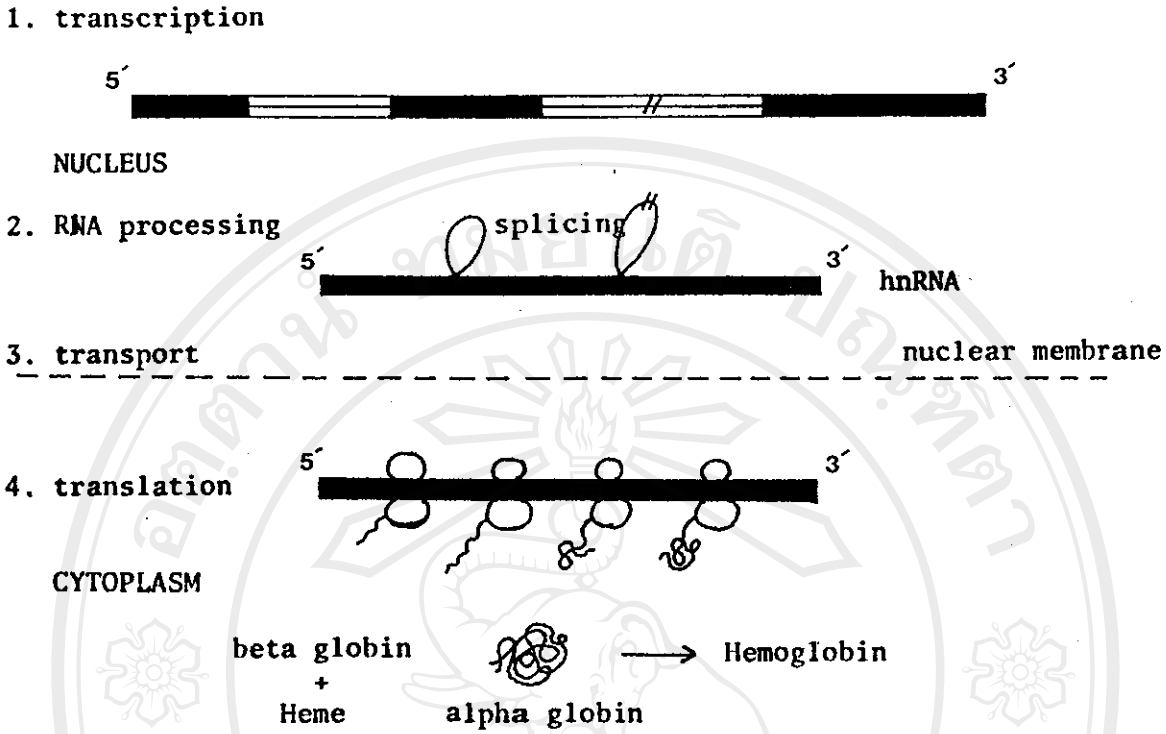

Figure I.4 Diagram of globin chain synthesis.

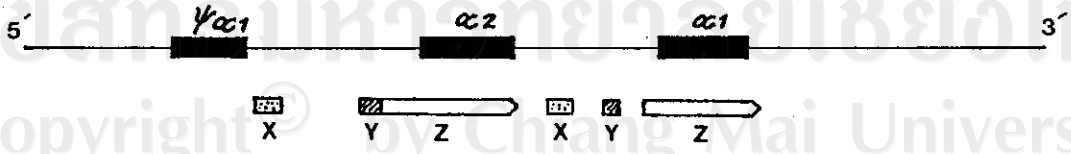

Figure I.6 Details of the alpha globin duplication unit.

X,Y,Z = the homology blocks.

I.5 Alpha-thalassemia :

The most common mechanism that produces alpha-thalassemia is deletion of one alpha globin gene locus or both loci. In some alpha-thalassemias, the impairment of alpha globin gene function is not the result of deletion. These are classified as the nondeletional alpha-thalassemias (16).

I.5.1 The deletional types :

Several distinct clinical forms of alpha-thalassemia are recognized. Table I.1 represents these forms in order of increasing severity. It is assumed that the normal condition of the alpha globin gene is for two pairs of loci to be functional ($\alpha\alpha/\alpha\alpha$), and that the severity of alpha-thalassemias depend on how many of alpha globin gene loci have been deleted (8).

Table I.1: Alpha-thalassemias in order of increasing severity.

no of α-gene loci deletion	types	Clinical feature
1 α-gene loci deleted ($-\alpha/\alpha\alpha$)	α-thal-2	none
2 α-gene loci deleted ($--/\alpha\alpha$)	α-thal-1	mild anemia
3 α-gene loci deleted ($--/-\alpha$)	Hb H disease	moderate to severe anemia, icterus, and spleenomegaly.
4 α-gene loci deleted ($--/--$)	Hb Bart's hydrops fetalis	

I.5.1.1 The Alpha-thalassemia-2 (-α/haplotype) :

Alpha-thalassemia-2 is the result of a deletion of one alpha-globin gene locus (11, 17). Two different arrangements have been observed (Figure I.5)

A. Leftward type ($-\alpha^{4.2}$) This type involves a deletion of 4.2 kb of DNA including the $alpha_2$ globin gene, it is common in Oriental population (17,21).

B. Rightward type ($-\alpha^{3.7}$) This type involves a deletion of 3.7 kb of DNA between the $alpha_1$ and $alpha_2$ globin gene, and some part of both genes (18).This type is predominated in all ethnic groups where alpha-thalassemia is present, including Thailand. (19,20).

The alpha globin genes are embeded within two homologous segments of DNA each approximately 4 kb long. Insertions or deletions have further divided each of these regions into three homologous subsegments (X,Y and Z as Figure I.6) (22). The Z $alpha_1$ and Z $alpha_2$ are highly conserved because of conversion and crossover fixation, which is called concerted evolution. Unequal crossover between the $alpha_1$ and $alpha_2$ globin genes serves as a mechanism for generating changes in gene number. The result of this event will be a chromosome bearing a single alpha globin gene, with the other chromosome having three alpha globin gene loci (9). The ($-\alpha^{4.2}$) is the result of a crossover at the X-block 3' to the $alpha_1$ gene on one chromosome with the block 3' the $alpha_2$ gene of the other, the ($-\alpha^{3.7}$) defect results

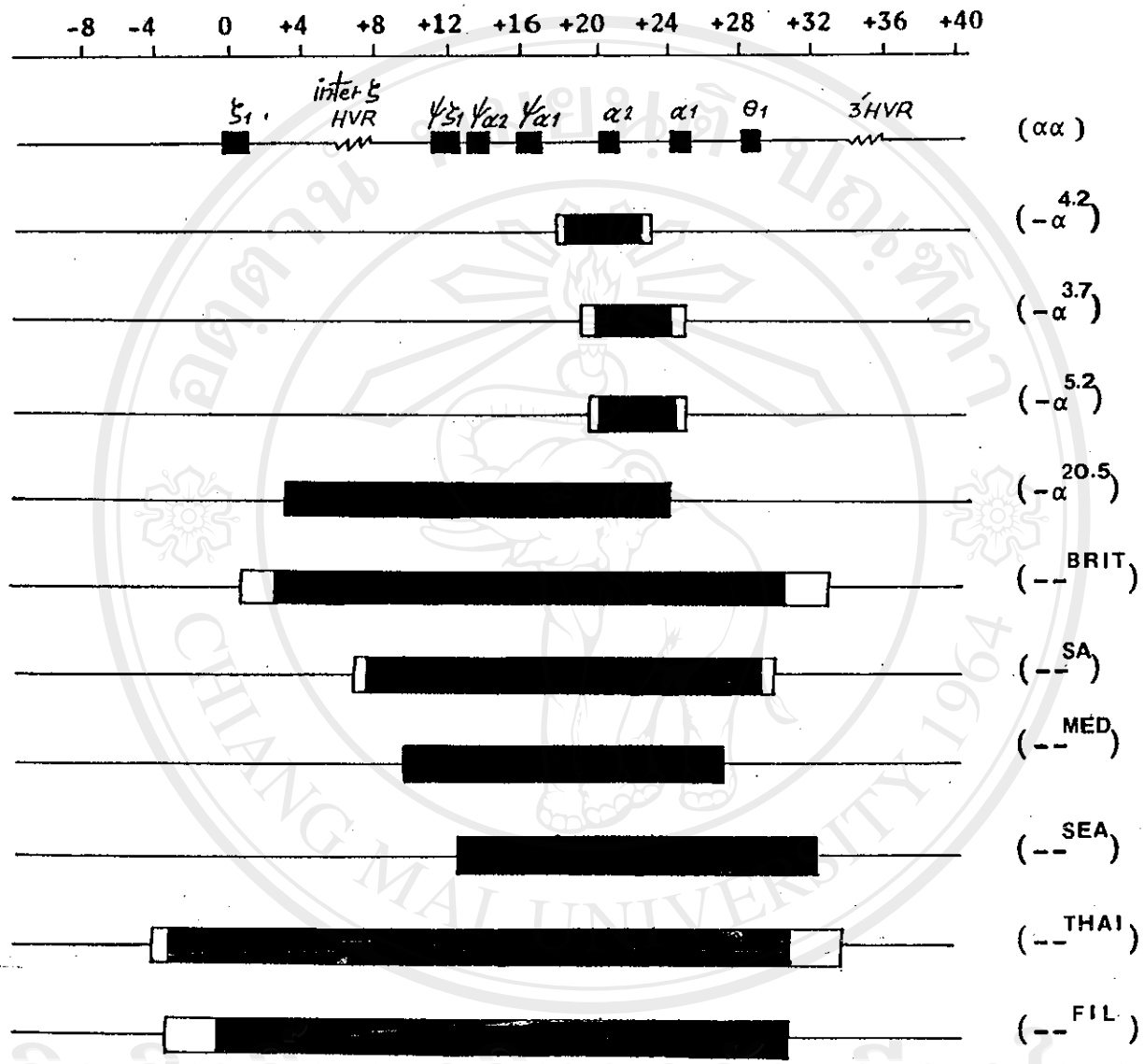

Figure I.5 Deletional mapping of the human alpha globin gene cluster.

Coordinations are given in kb ; 0 represents the Zeta-mRNA cap site.

from inter- or intrachromosomal recombination between two homologous Z-segments (Figure I.7)

Restriction mapping and DNA sequence analysis of the mutants indicate that different($-\alpha^{3.7}$) chromosomes are the result of at least three independent events that distinguish three ($-\alpha^{3.7}$) subtypes, charaecterized by different crossover sites between the misaligned $alpha_2$ and $alpha_1$ gene (23) (Figure I.8).

Subtype I : result from a crossover before a 7 bp insert, which is present at the 3 end of IVS-2 of the $alpha_1$ gene.

Subtype II : results from a crossover in the third exon of the $alpha_1$ gene (3'coding region).

Subtype III : results from a crossover in the 3'non coding region of the $alpha_1$ gene.

In most populations, the crossovers are in region I and II (subtype I and II). Only in Melanesia, the crossover in region III (subtype III) is the most prevalent ($-\alpha^{3.7}$) defect.These observations could be interpreted as a reflection of the rate at which the various crossovers occur in the large (region I = 1,436 bp), the medium (region II = 171 bp), or the small regions (region III = 46 bp) of homology between the two alpha globin genes.

Two other rare alpha-thalassemia mutants : result from the deletion of one alpha globin gene (24).

C. ($-\alpha^{5.2}$) This rare type involves a deletion of 5.2 kb of DNA between both alpha globin genes and some parts of them as in the case of ($-\alpha^{3.7}$), but the deletional fragment is larger.

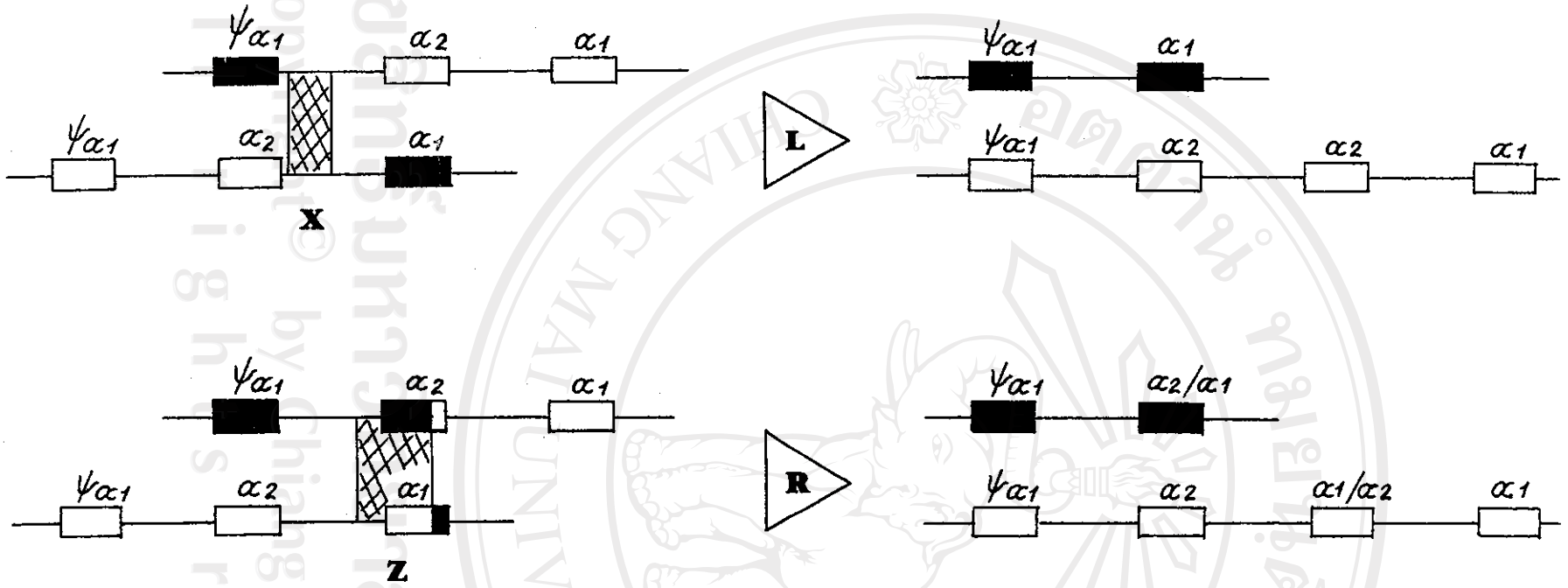

Figure I.7 The alpha$_1$ and alpha$_2$ globin genes in the positions assumed during a misaligned crossover.

L = leftward alpha-thal-2 : crossingover in X-block.

R = rightward alpha-thal-2 : crossingover in Z-block.

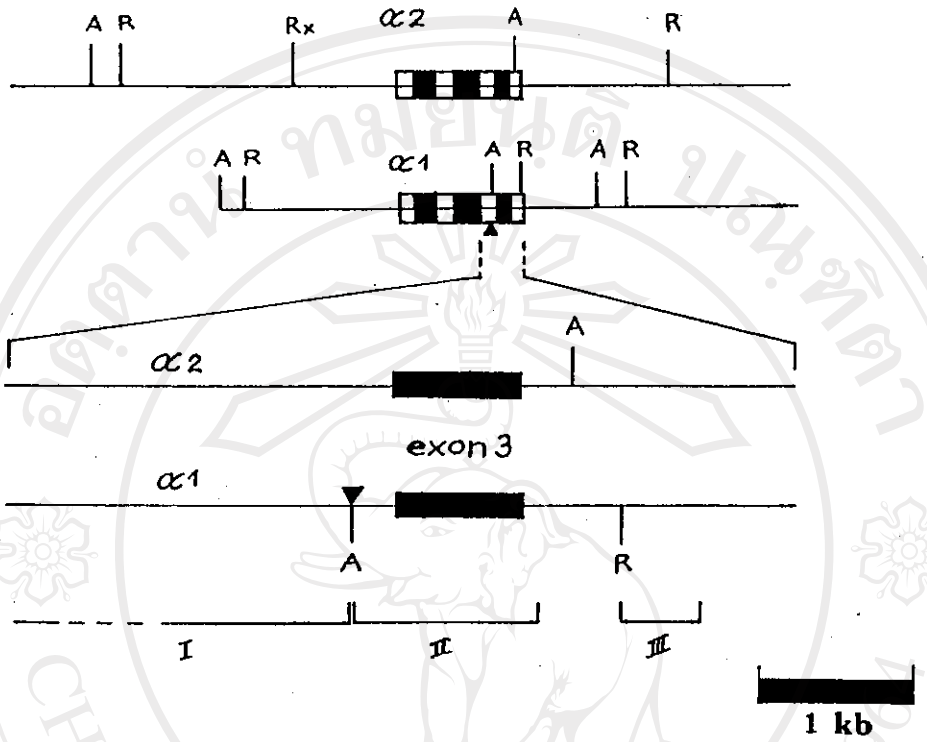

Figure I.8 Detailed comparison of the alpha$_1$ and alpha$_2$ genes beyond the 7 bp insert (▲) in alpha$_1$ globin gene. Restriction site A (Apa I), R (Rsa I) are indicated. Segments I, II, and III correspond to the rightward alpha-thal-2 crossover regions.

D.($-\alpha^{20.5}$) This type involves a large deletion of DNA about 20.5 kb including the interzeta hypervariable region (HVR) up to the $alpha_1$ gene (Figure I.5).

I.5.1.2. Alpha-thalassemia-1 (--/haplotype) :

Alpha-thalassemia-1 occurs from a deletion of both alpha globin genes loci. Several arrangements have been observed (Figure I.5) :

A.($--^{BRIT}$) This rare mutant involves a large deletion including the interzeta-HVR up to theta gene (θ).

B.($--^{SA}$) This rare mutant involves a large deletion as ($--^{BRIT}$) but shorter at the 5'breakpoint.

C.($--^{MED}$) This type involves a large deletion, the 5'breakpoint lies just upstream of the pseudozeta gene. The 3'breakpoint is at coordinate +28. This type is most common in the Mediteranean region.

D.($--^{SEA}$) This type involves a large deletion, the 5'breakpoint lies in the third exon of pseudozeta gene. The 3'breakpoint is at coordinate +33. This type is common in the Southeast Asians including Thais (25). The deletion at the 5'end is polymorphic, and thus provides a useful genetic marker for the alpha globin gene complex. By using the restriction enzyme Bgl II together with Bam HI and hybridization with a zeta probe, it is possible to identify at least three different haplotypes in Thai population, on the basis of the variable length of the interzeta-HVR 10.5, 11.3 and 12.0 kb (19).

E. $(--^{THAI})$ This rare type involves a deletion of the entire zeta-alpha-complex. The 5 breakpoint lies between corrdinates -3 to -4 and the 3' breakpoint lies between coordinates +31 to +34. This type was found in Thai subjects (25).

F. $(--^{FIL})$ This rare type involves a large deletion as $(--^{THAI})$ but the position of the breakpoint is different. The 5 breakpoint lies between coordinates 0 and -3, the 3' breakpoint lies close to coordinate +31. This type was found in Filipino subjects (25).

I.5.1.3. Hb H disease $(--/-\alpha)$:

Hb H disease occurs most frequently in Southeast Asian and more rarely in Mediteranean populations. This disorder commonly results from double heterozygosity for alpha-thalassemia-1 and alpha-thalassemia-2 (26). Howerver, it may also be due to from alpha-thalassemia-1 and a nondeletional alpha-thalassemia gene or to the homozygous state of nondeletional alpha-thalassemia genes such as hemoglobin Constant Spring (1).

I.5.1.4. Hemoglobin Bart's hydrops fetalis :

Hemoglobin Bart's hydrops fetalis occurs frequently in South east Asia and in some of the Mediteranean Island population (1). This disorder is the expression of the homozygous state of alpha-thalassemia-1 genes. Gene mapping studies have demonstrated the

deletion of four alpha globin genes (--/--) that cause the absence of alpha chain synthesis.

I.5.2 The nondeletional types :

The nondeletional alpha-thalassemias are characterized by decreased the alpha globin synthesis in the presence of two intact alpha globin genes (1,27). These comprise a distinct and apparently heterogeneous group of mutations as shown in Table I.2

Table I.2 : Alpha-thalassemia mutants.

Molecular mechanism of thalassemia effect	Structural abnormality in alpha globin chain
(1) Loss of normal terminator	Elongated globin chain : Hb Constant Spring (142) ter→Gln : Hb Icaria (142) ter→Lys : Hb Koya Dora (124) ter→Ser : Hb Seal Rock (142) ter→Glu
(2) Globin or hemoglobin instability	: Hb Quong Sze (125)Leu→Pro

(3) Unknown mechanism	: Hb Suan Dok
(Postulated post-translational	(109) Leu→Arg
instability)	: Hb Petha Tikvah
	(110) Ala→Asp

The normal termination codon for both alpha and beta globin is UAA : alpha globin mRNA contains 109 nucleotides in the 3' noncoding region. Replacement of a single nucleotide in the normal terminator to give a codon that specifics an amino acid rather than termination leads to synthesis of elongated globin chains, the resulting mutant hemoglobins are structurally different only in the amino acid at the terminator codon position as indicated in Table I.2.The first of these hemoglobins to be discovered was hemoglobin Constant Spring (Hb CS). The Hb CS mutant occurs in the $alpha_2$ gene, Hb CS $alpha_2$ mRNA is read to the next in phase terminator to give a globin with 156 to 172 amino acids rather than the normal 141 amino acids (28). Hb CS is commonly found in Southeast Asia including Thailand. It has been observed occasionally in Mediteranean population. This alpha chain variant which is produced at a markedly reduced rate and hence gives rise to the phenotype of alpha-thalassemia-2. Therefore, compound heterozygosity of alpha-thalassemia-1 and Hb CS results in Hb H disease as described above, more specifically Hb CS-Hb H disease (19).